1 Networks

If G is a graph (a collection of nodes, with (directed) edges between some of the nodes), we associated to it an *incidence matrix* A. This matrix has one column for each node, and one row for each edge. For each edge in the graph, insert a row with a -1 in the column corresponding to the starting node, and $+1$ in the column corresponding to the ending node. The cool thing is that the various fundamental subspaces of A tell us things about graph G. A combination of rows which is 0 corresponds to a loop in the graph – thus loops come from vectors in the left nullspace of A. The nullspace of A consists of the single vector $(1, \ldots, 1)$.

2 Orthogonality

Two vectors are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. If V is a subspace, then the set of all vectors orthogonal to every vector $\mathbf{v}$ in V is a subspace too: this is the orthogonal complement of V, denoted $V^\perp$.

A matrix Q is said to orthogonal if its columns are an orthonormal set of vectors, meaning that $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ if $i \neq j$, where $\mathbf{v}_i \cdot \mathbf{v}_i = 1$ for every i. This implies that $Q^T Q = I$. Orthogonal matrices also satisfy $|Q\mathbf{x}| = |\mathbf{x}|$ for any vector $\mathbf{x}$.

Another useful property is that if you have an orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_n$ for a space, then for any vector, $\mathbf{x} = (\mathbf{x} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{x} \cdot \mathbf{q}_n)\mathbf{q}_n$. When your set is just $\mathbf{e}_1, \ldots, \mathbf{e}_n$ this just corresponds to writing out a vector as the sum of its components in the various coordinate directions. If $\mathbf{q}_1, \ldots, \mathbf{q}_k$ is a set of orthonormal vectors that isn't a basis (because they don't span the whole space), then $\mathbf{x} = (\mathbf{x} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{x} \cdot \mathbf{q}_k)\mathbf{q}_k$ is the projection of $\mathbf{x}$ onto the span of $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

3 Projection matrices and least squares

Suppose that V is a subspace of $\mathbb{R}^n$. If $\mathbf{x}$ is a vector which isn't contained in V, it's often useful to know the point of V that's the closest to $\mathbf{x}$. It turns out that there's a matrix P such that for any vector $\mathbf{x}$, the closest point in V to $\mathbf{x}$ is just $P\mathbf{x}$. P is the called a projection matrix. To get the matrix, you need to find a basis for V, form a matrix A by putting all your basis vectors together as the columns, and then compute $P = A(A^T A)^{-1} A^T$ (a formula well worth remembering). Note that $A^T A$ is a square matrix, so the inverse makes sense.

Two basic properties of projection matrices: $P^2 = P$ and $P^T = P$. If you choose your matrix $A = Q$ to be orthogonal (i.e. choose an orthonormal basis for V), this takes a particularly simple form: $P = Q(Q^T Q)^{-1} Q^T = Q(I)Q^T = QQ^T$.

Suppose we're trying to solve $A\mathbf{x} = \mathbf{b}$. If $\mathbf{b}$ isn't in the column space of A, there can't be any solutions. The best we can do is to find the point $\mathbf{p}$ in $V = \text{col}(A)$ closest to $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ does have a solution. A bit of work (explained in the book) shows that the closest $\mathbf{p}$ is the projection of $\mathbf{b}$ onto the column space of A: $\mathbf{p} = A(A^T A)^{-1} A^T \mathbf{b}$. Then we need to solve $A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}$. Applying A^T to both sides we need to solve $A^T A\mathbf{x} = A^T \mathbf{b}$. This is a general guideline: if you want to solve $A\mathbf{x} = \mathbf{b}$, but there's no solution, the best-fit approximation will be the solution to $A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$; this means $\mathbf{p} = A\hat{\mathbf{x}}$. The error $\mathbf{e} = \mathbf{b} - \mathbf{p}$ will always be orthogonal to V.

A typical setting where this method comes in handy is in trying to fit a curve of a specific type through a bunch of data points: say we have a bunch of pairs (x_1, y_1), $\ldots$, (x_n, y_n) and want to find the best-fit parabola $y = at^2 + bt + c$. To find a parabola going through all the points, we'd want a, b, and c to satisfy $y_i = ax_i^2 + bx_i + c$ for every i. This is a linear equation ("$A\mathbf{x} = \mathbf{b}$") in a, b, c:

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ & \cdots & \\ x_n^2 & x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

In general this is lots of equations in three variables, and there are no solutions (the right side isn't in the column space of the matrix). That means there aren't any parabolas through all the points. But you can solve $A^T A\mathbf{x} = A^T \mathbf{b}$ instead, and that will be the best fit.

4 Gram-Schmidt and $A = QR$

Say we're handed a list of independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The Gram-Schmidt process will start with these and then produce a new set vectors with the same span, but which are orthonormal. This is convenient, since dealing with projection matrices and many other things is easier when you have an orthonormal basis. Here's how it works: the first vector $\mathbf{w}_1$ is just the original vector you started with. To get $\mathbf{w}_2$, take $\mathbf{v}_2$, and subtract its component in the $\mathbf{w}_1$ direction: $\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} \mathbf{w}_1$. Similarly, $\mathbf{w}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} \mathbf{w}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2} \mathbf{w}_2$. This produces a set of orthogonal vectors. At the end, you want to normalize by dividing each of your vectors by its length.

If you imagine putting your vectors as the columns of a matrix, and then carrying out Gram-Schmidt by doing basic column operations, each step corresponds to multiplying A on the right (since we're doing column instead of row operations). Combining all the elimination steps, in the end you get $AS = Q$, where S is upper triangular and Q is orthogonal. Multiply both sides by S^{-1} on the right, and we get $A = QS^{-1} = QR$.

Keeping track of all the elimination matrices can get annoying, so here's a shortcut. First just compute Q via Gram-Schmidt as usual. We know that $Q^TA = Q^TQR = R$, since Q is orthogonal. If you don't want to keep track of elimination matrices at every step, you can just compute Q first by Gram-Schmidt, and then get R using $R = Q^TA$.

5 Determinants

A determinant takes a square matrix and outputs a number. Be sure you know the following useful properties of this operation:

- Determinant is linear in each row and column.
- Determinant is unchanged when adding one row/column to another row/column.
- Determinant changes sign when swapping two rows/columns.
- $\det A = 0$ if and only if A has a nonzero nullspace (i.e. it's not invertible).
- The determinant of A is equal to the volume of the parallelpiped defined by the columns of A; this shows up as the Jacobian in change of coordinates for 2- or 3-dimensional integrals.

We have three main methods for computing determinants. I'm not going to explain how each goes, but make sure you know them all. Here they are, with some suggestions one when each one is useful.

- Cofactor expansion. Especially suitable when the matrix has a row or column with only a couple nonzero entries. Also good in a lot of these induction-type problems where you want to know the determinants of a sequence of $n \times n$ matrices of similar shape.
- The big formula. Good if there are only a couple nonzero terms, or if the matrix is 2×2 or 3×3 (in which case the big formula isn't so big after all, having only 2 or 6 terms).
- Product of pivots. If a matrix doesn't fit any of the above (i.e. it's at least 4×4 and doesn't have many zeroes), this is probably the way to go.

Perhaps the following should qualify as a basic methods:

- Triangular matrices: just multiply the diagonal entries.
- Reduce to simpler matrix by row/column operations. These don't change the determinant, so if you can do something sneaky like reduce to an triangular matrix or a matrix with lots of 0's using just a couple row/column operations (including row swaps), you're in business.

6 Cramer's rule

Cramer's rule is a way to solve linear systems $A\mathbf{x} = \mathbf{b}$, where A is square. Here's what you do. Let B_i be the matrix obtained by replacing the i^{th} column of A with the vector $\mathbf{b}$. The $x_i = \det B_i / \det A$. As a consequence, we get a formula for A^{-1}: we have $A^{-1} = C^T / \det A$, where C is the matrix of cofactors of A. These are not efficient ways to solve $A\mathbf{x} = \mathbf{b}$ in general, since you have to take a lot of determinant. Nevertheless it can sometimes be very helpful to have an explicit formula like this.

7 Eigenvalues and eigenvectors

A vector $\mathbf{x}$ is called an eigenvector of the matrix A if $A\mathbf{x} = \lambda\mathbf{x}$, where λ is a number (called the eigenvalue): this means that $A\mathbf{x}$ points in the same direction as $\mathbf{x}$ (or the opposite direction, in case $\lambda < 0$).

This is equivalent to saying that $(A - \lambda I)\mathbf{x} = \vec{0}$, which is to say that $\mathbf{x}$ is in the nullspace of $A - \lambda I$. This is the observation that lets us find the eigenvectors. For most values of λ, the matrix $A - \lambda I$ won't have a nullspace at all. The only times it does is when $\det(A - \lambda I) = 0$, and so the eigenvalues are precisely the solutions of $\det(A - \lambda I) = 0$. Once you've find an eigenvalue, the way to find the corresponding eigenvector is to write down the matrix $A - \lambda I$ for that value of λ, and then find something in the nullspace using the usual procedure (elimination, special solutions).

There are a couple cases where this has particular geometric significance: $\mathbf{x}$ is an eigenvector with $\lambda = 1$ means that $\mathbf{x}$ doesn't change when you apply A; with $\lambda = 0$ means that $A\mathbf{x} = \vec{0}$, i.e. $\mathbf{x}$ is in the nullspace of A; with $\lambda = -1$ means that the direction of $\mathbf{x}$ is reversed when we apply A to $\mathbf{x}$.

The sum of the eigenvectors is equal to the trace, which is defined to be the sum of the diagonal entries. The product of the eigenvectors is equal to the determinant. This can be useful either for finding the determinant/trace if you somehow know all the eigenvalues, or finding one eigenvalue if you know the determinant/trace and all of the other eigenvalues.